Politecnico di Milano
Scuola di Ingegneria Industriale e dell'Informazione
Applied Statistics
June 15th, 2020

Problem n.4

The file `revenues.txt` collects the average daily revenues y [k€] during the lockdown of 70 minimarkets located in Milan. The dataset also reports the UTM coordinates s_i of the shops, the resident population in the neighborhood around the shop $p(s_i)$, and the Euclidean distance $d(s_i)$ [m] between the location of the shop and the Duomo $d(s_i) = \|s_i - s_d\|$, with $s_d = (514711.6, 5033903.0)$. Consider for the revenue $y(s_i)$, $i = 1, ...70$, the following model

$$y(s_i) = a_0 + a_1 \cdot p(s_i) + \delta(s_i),$$

with $\delta(s_i)$ a stationary residual.

a) Estimate via generalized least squares the parameters a_0, a_1 of the model. Report the model estimated for $\delta(s_i)$, and discuss the model assumptions.

b) Provide a kriging prediction $y^*(s_0)$ of the revenues at a shop located in the Brera district at location $s_0 = (514703.8, 5035569.3)$. For this purpose, use a point estimate of the resident population $p(s_0)$ obtained through a linear model in the variable *distance from the Duomo* (detail the model assumptions for $p(s_0)$ and its point estimate).

c) Report the kriging variance $\sigma^2(s_0)$ of the point prediction at point (b). Would you deem the variance $\sigma^2(s_0)$ to be fully representative of the uncertainty associated with the prediction $y^*(s_0)$?